Heriot-Watt University
Research Gateway

Adversarial Approach to Prediction in Atari Games

Citation for published version:
Andrecki, M & Taylor, NK 2017, 'Adversarial Approach to Prediction in Atari Games', Paper presented at 2017 EPSRC CDT Student Conference – Oxford, Bristol and Edinburgh, Oxford, United Kingdom, 7/06/17 - 7/06/17.

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. May. 2023

Adversarial Approach to Prediction in Atari Games

M. Andrecki, N.K. Taylor

M.Andrecki@hw.ac.uk

Abstract

Recent advances in neural networks have resulted in reliable predictions of observations in Atari games. The quality of predictions can be improved by addition of a generative adversarial approach to the training. This research investigates this development and the benefits of unsupervised predictive learning for reinforcement learning agents.

I. Introduction

Reinforcement learning (RL) based on artificial neural networks (ANN) has seen great successes in recent years. A notable recent breakthrough came from [6], where an artificial agent learnt to play games from raw pixels on a screen along with scores produced by Atari game console simulation. The algorithm managed to surpass human performance on many classic 1980s games. However, RL has not yet achieved similar successes for real world tasks, such as controlling robotic manipulation.

RL requires dozens of hours of gameplay in order to perform at human level. This training time is currently a significant obstacle outside simulation systems. It has been argued that pure RL is data inefficient because the reward signal – the only feedback used – is sparse and contains little information.

To combat this, various forms of unsupervised learning have been added to the pure task of reward maximisation. Similar capabilities can be applied to estimating the value of a given screen frame (usual RL) and for prediction of rewards or pixel values in subsequent timesteps. This idea was the basis for [4], where an ANN implementing an agent attempted to solve multiple related unsupervised and reinforced tasks at the same time.

This paper investigates how training an ANN for prediction of future observations can improve data efficiency in RL problems. In particular, in the relatively simple (but high dimensional) setting of Atari game playing.

II. Predictive learning

The ability to predict the future state of the world can speed up RL in many ways.

Value learning: If an agent knows what happens after a given action is taken, it can update its value function estimate based on internal simulation rather than through experience. This is especially useful for robots – for which taking actions in the real world is very time consuming (relative to simulation). This idea is captured by the Predictron algorithm presented in [10].

State representation: If a given state representation allows for reliable prediction of future percepts, it is a sound basis for value function approximation. Attempting to enforce a useful state representation in an ANN can be seen in [4, 5].

Prediction of the results of an action can be used to **select an action** which is more likely to bring a desirable outcome.

In addition to this, a learnt model can be examined to explain an agent's decisions or performance. [7, 9] are examples of attempts at understanding an ANN's representation.

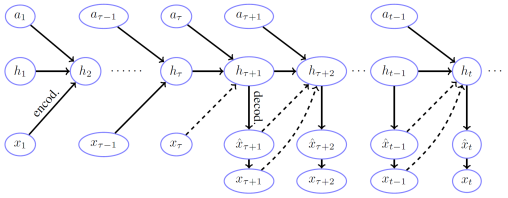

Figure 1: *Architecture used in [1]. a is a joystick action, x is frame pixel values, $\hat{x}$ is frame reconstruction, h is a representation of game state – implemented with an LSTM – computed with use of frame pixels, last action and previous state.*

III. Predicting in Atari worlds

Over the last year ANN-based predictions of Atari game frames reached an impressive level. First [8] demonstrated an ability to simulate deterministic games reliably for hundreds of frames ahead. Then [1] improved on their approach and increased the quality of predictions.

Figure 1 captures dependencies of different parts of a neural architecture used by [1]. Screen pixel dimensionality is reduced by passing it through convolutional layers. Then predictions are made in the resulting reduced space using an LSTM [3]. Afterwards, a frame is generated using deconvolution. The objective function is the mean squared error between the predicted and the actual frame.

Overall, the predictive models function exceptionally well. However, they are restricted to deterministic environments. At times the generated images are blurred and small objects (e.g. projectiles) can be missing.

IV. Proposed modifications

This research attempts to enhance the [1] method so that it can cope with mild stochasticity. The resulting model will be examined to understand how the networks choose to represent different aspects of the environment, like uncertainty about object positions or unobservable states, e.g. object velocity.

The quality of predictions could be improved with the use of Generative Adversarial Networks (GANs) [2]. In this scheme two networks are trained in parallel. One learns to generate fake data (e.g. images), and the other learns to distinguish between real and generated data. GANs have proved a strong method for sampling from complex distributions, such as natural images.

In a situation where the physical process is stochastic but the network is forced to make a single valid prediction, a conventional ANN will output an average of the possible results. Thus uncertainty in the position of some objects translates into a blurred image. GANs are incentivised to generate real-looking images – a blurry image would be immediately recognised as a fake and thus it will not be output.

Additionally, knowledge captured by the network will be inspected. For example, what partial subset of a state has to be known to predict the next state?

Finally, the utility of the resulting model will be tested from an RL perspective in ways previously described.

References

[1] S. Chiappa et al. Recurrent environment simulators. 2017.

[2] I. Goodfellow et al. Generative adversarial nets. pages 2672–2680, 2014.

[3] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[4] M. Jaderberg et al. Reinforcement learning with unsupervised auxiliary tasks. *CoRR*, abs/1611.05397, 2016.

[5] X. Li et al. Recurrent reinforcement learning: A hybrid approach. *CoRR*, abs/1509.03044, 2015.

[6] Mnih et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[7] Mordvintsev et al. Inceptionism: Going deeper into neural networks, june 2015. *URL http://googleresearch. blogspot. com/2015/06/inceptionism-going-deeper-into-neural. html*.

[8] J. Oh et al. Action-conditional video prediction using deep networks in atari games. *CoRR*, abs/1507.08750, 2015.

[9] A. Radford et al. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.

[10] D. Silver et al. The predictron: End-to-end learning and planning. *CoRR*, abs/1612.08810, 2016.